Subsign Detection and Classification System for Automated Traffic-sign Inventory Systems

Lykele Hazelhoff[1,2], Ron op het Veld[2], Ivo Creusen[1,2] and Peter H. N. de With[1,2]
[1]*CycloMedia Technology B.V, Zaltbommel, The Netherlands*
[2]*Eindhoven University of Technology, Eindhoven, The Netherlands*

Keywords: Object Detection, Traffic Sign Recognition, Object Classification.

Abstract: Road safety is influenced by the accurate placement and visibility of road signs, which are maintained based on inventories of traffic signs. These inventories are created (semi-)automatically from street-level images, based on object detection and classification. These systems often neglect the present complimentary signs (subsigns), although clearly important for the meaning and validity of signs. This paper presents a generic, learning-based approach for both detection and classification of subsigns, which is based on the same principles as the system employed for finding traffic signs and can be used as an extension to automated inventory systems. The system starts with detection of subsigns in a region below each detected sign, followed by analysis of the results obtained for all capturings of the same sign. When a subsign is found, the corresponding pixel regions are extracted and subject to classification. This recognition system is evaluated on 3,104 signs (397 with subsign) identified by an existing inventory system. At a detection rate of 98%, only 757 signs (24.4% of the signs) are labeled as containing a subsign, while 91.4% of the subsigns of a class known to our classifier are also classified correctly.

1 INTRODUCTION

Road safety is strongly influenced by the correct placement and accurate visibility of traffic signs to e.g. warn road users for upcoming dangerous situations or inform them about speed limits or other restrictions. The validity and legal meaning of these road signs are commonly affected by co-attached complimentary signs, which contain e.g. directions, time restrictions or arbitrary texts. Figure 1 displays examples of such complimentary signs. As the visibility of road signs degrades over time due to aging, accidents or vandalism, accurate inventories of road signs are of significant interest to governmental instances and subcontractors tasked with road maintenance. Moreover, these inventories are applicable to driver assistance systems or autonomous vehicles.

These inventories can be generated manually, tracking all roads, but efficiency can be improved by employing street-level images, together with object detection and classification techniques to retrieve the traffic-sign positions and types. Multiple of such systems exists , i.e.(Hazelhoff et al., 2012),(Maldonado-Bascon et al., 2007),(Maldonado-Bascon et al., 2008),(Overett and Petersson, 2011),(Timofte et al., 2009),(Timofte et al., 2011). These generally start by processing every image to detect the present road signs, which are then often tracked over multiple consecutive frames. Classification of the signs is usually performed either directly after detection, or after tracking. Recognition scores over 90% are reported, where some systems include over 90 different sign types.

The reported systems focus primarily on the recognition of road signs, but commonly ignore the presence of complimentary signs. However, these subsigns are of great importance during analysis of the inventory results, and are therefore typically added manually. This is a time-consuming process, especially since subsigns only occur for the minor-

(a) (b) (c) (d)

Figure 1: Example of a complimentary signs altering the meaning of road signs.

Hazelhoff L., op het Veld R., Creusen I. and H. N. de With P..
Subsign Detection and Classification System for Automated Traffic-sign Inventory Systems.
DOI: 10.5220/0004654402620268
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 262-268
ISBN: 978-989-758-004-8

ity of signs. Extension of these automated traffic-sign inventory systems with a subsign recognition module would increase the inventory generation efficiency, and would decrease the required manual interaction. However, subsign detection and recognition is rather difficult, as the subsign contents vary greatly and sometimes contain arbitrary texts and/or custom symbols. Furthermore, these signs are smaller than normal road signs and consist of less discriminative colors, as they are usually white. Also, the capturing quality and conditions are varying, as the capturings are made outdoors during different weather conditions and from driving vehicles, typically with large inter-capturing distances at higher driving speeds.

In literature, few publications report on the recognition of subsigns. In (Hamdoun et al., 2008), rectangle detection is employed in a region below road signs to retrieve present supplementary signs, followed by a classification stage to solely retrieve exit-lane subsigns. Rectangle detection is also exploited in (Nienhuser et al., 2010), where found rectangles are classified using a two-stage cascade, aiming at discrimination between both subsign and non-subsign rectangles and between 4 different subsign types. In (Python et al., 2012), a region growing approach is described and compared against several other techniques, where the proposed method achieves a correct detection rate over 70%.

This paper describes a generic and learning-based approach for both detection and classification of subsigns. The work forms an extension to our existing traffic-sign inventory system (Hazelhoff et al., 2012), but due to the generic nature of our algorithm, it is also applicable to other, similar systems such as (Maldonado-Bascon et al., 2007), (Maldonado-Bascon et al., 2008), (Overett and Petersson, 2011), (Timofte et al., 2009) and (Timofte et al., 2011)). Instead of treating the complimentary signs as an additional sign class, and thereby searching the complete image for complimentary signs, the output of existing road-sign inventory systems, like any of the abovementioned systems, is exploited. This narrows the search area to the regions below the identified traffic signs, which increases robustness, since objects with a similar appearance compared to subsigns (i.e. white rectangles) occur frequently in real-world situations. The subsign recognition system exploits both the single-image detections and tracked detections (in this paper referred to as 3*D* signs) given by existing inventory systems.

The system starts by detection of subsigns in a fixed region below each of the detected signs, as the vast majority of subsigns are located below road signs. Afterwards, the subsign detection results obtained for each detection of a 3*D* sign, are combined to improve robustness. When a subsign is found, the corresponding pixel regions are extracted, which are then subject to classification, to retrieve either the subsign type or a subsign-with-text code. This system is evaluated on a large, real-world dataset containing 3,104 signs (397 signs with subsign), with 29 different subsign types for classification. It will be shown that subsign detection is indeed possible with reasonable performance, even with a generic concept.

(a) (b) (c) (d) (e)

Figure 2: Example of a 3*D* sign, consisting of multiple detections of the same traffic sign tracked over multiple consecutive capturings.

The remainder of this paper is organized as follows. Section 2 contains the system overview of our subsign detection and classification system, which is described in detail in Sect. 3. The performed experiments and results are found in Sect. 4, followed by the conclusions in Sect. 5.

2 SYSTEM OVERVIEW

The system for automatic recognition of subsigns described in this paper operates on 3*D* signs detected by our traffic-sign inventory system (Hazelhoff et al., 2012). These 3*D* signs consist of multiple detections of the same road sign, tracked over consecutive image frames. An example of an input 3*D* sign is shown in Fig. 2. The system overview of the subsign recognition system is depicted in Fig. 3, and the four primary modules are briefly described below.

1. *Single-image detection:* The region below each detection given by the inventory system is divided in overlapping windows. Each window is described based on densely extracted SIFT descriptors, which are subject to classification with a linear Support Vector Machine (SVM). The maximum SVM output of one out of all windows is returned for each analyzed detection.
2. *Multiview Detection:* The single-image detection results are combined to determine the presence of a subsign for each 3*D* sign.
3. *Subsign Localization:* When a subsign is found for the 3*D* sign, the pixel region corresponding to the subsign, is retrieved for each detection with a positive SVM output during the single-image detection stage.

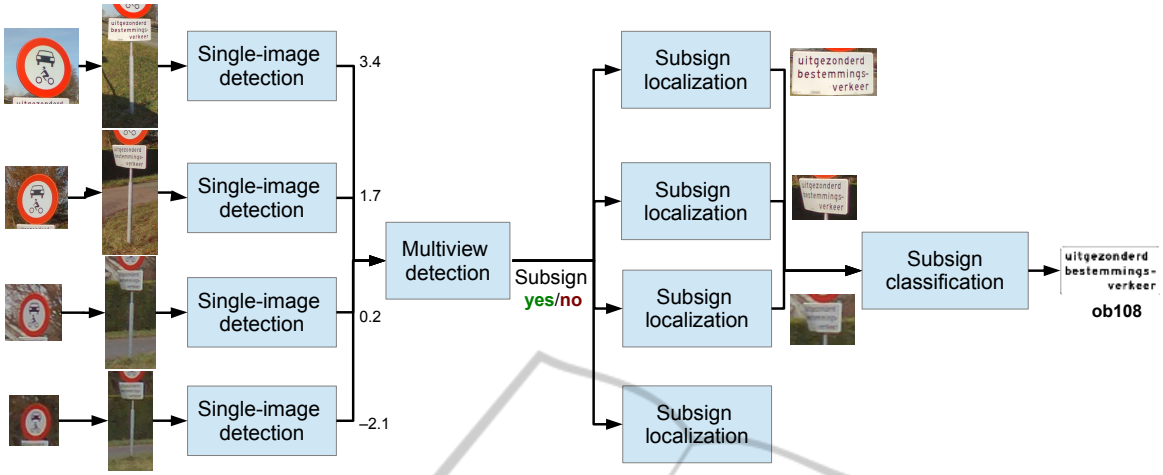

Figure 3: System overview of the subsign recognition system.

4. *Subsign Classification:* The identified subsign regions are subject to classification. Afterwards, the subsign type is retrieved based on weighted voting.

3 ALGORITHM DESCRIPTION

3.1 Single-image Subsign Detection

At first, each detection of the input $3D$ sign is analyzed independently, where a measure for the possible presence of a complimentary sign is computed. This stage starts by extraction of the region below the detection, where we employ a region height of twice the detection height, as subsigns are not necessarily located directly below the road signs. Within this region, a sliding-window approach is employed, with a windows size of 80% of the detection width and 30% of the detection height, which corresponds to the typical subsign size. Each extracted window is resized to a standard size of 120×45 pixels, and afterwards, SIFT descriptors (Lowe, 2004) are extracted from a dense grid at a single scale. These descriptors are selected since they are very robust against variations in lighting conditions and small object deformations, e.g. caused by subsign rotation and skewness, which occur commonly. After concatenation of these descriptors, the resulting feature vector is subject to classification using a linear SVM. This classifier is trained on a large training set containing over $30,000$ subsigns and non-subsigns windows. After all windows are analyzed, the maximum SVM output for the subsign class over all evaluated windows is selected as subsign presence indicator.

3.2 Multiview Subsign Detection

Traffic-sign inventory systems commonly identify the same road signs in multiple images, where the corresponding detections are tracked over the subsequent image frames. As all different detections contain the same sign, but are captured from different distances and from various viewpoints, the robustness of the subsign detection stage is improved by combining the subsign detection scores obtained for each of the individual detections. As complimentary signs may not be visible from all different viewpoints, and may be visible more clearly from a close distance, we compare three different combination methods against each other. The first method averages the individual detection scores, the second selects the best score (i.e. the most likely subsign detection) and the third employs the median score. These methods are defined as:

- Average:
  $d_{3D,av} = mean(d_{det\,1}, d_{det\,2}, ..., d_{detN})$,
- Max:
  $d_{3D,ma} = max(d_{det\,1}, d_{det\,2}, ..., d_{detN})$,
- Median:
  $d_{3D,me} = median(d_{det\,1}, d_{det\,2}, ..., d_{detN})$.

In these formulas, d_{3D} denotes the corresponding detection score for the $3D$ traffic sign, and $d_{det\,i}$ denotes the detection score for the i-th detection, with $i \in (1, ..., N)$ and N being the total amount of detections of the respective $3D$ sign.

3.3 Subsign Localization

When the presence of a subsign is indicated for a $3D$ sign, the subsign pixel region is identified for all detections for which the single-image detection phase

results in a positive SVM output for the subsign class. During detection, the window with the highest SVM output is selected, whereas in this stage, this window is extended by all adjacent windows which also have a positive SVM output. This involves iterative selection of all windows which are overlapped by the current subsign region for at least 40%, where initially the window with the highest SVM output forms the current subsign region. Figure 4 portrays several examples of obtained subsign regions.

Figure 4: Examples of localized subsigns. This figure also displays examples of subsign classes contained in the sets employed for training of the classification stage.

3.4 Subsign Classification

Each localized subsign region is classified, where a standard object recognition method is reused for the new purpose, which essentially operates as follows. Each of the retrieved subsign regions is first resized to a predefined size, after which SIFT descriptors (Lowe, 2004) are extracted from a dense grid. These descriptors are then concatenated, and the resulting feature vector is $L2$ normalized afterwards. Then, multi-class classification is performed using linear SVMs in a One-versus-All setup. Subsequently, the classification result for the $3D$ sign is retrieved based on weighted voting, where the weight is defined as the difference in SVM outputs of the winning class and second class.

This classification system is trained on the detection output from over $100,000$ $3D$ signs, which are manually annotated. During this process, non-ideal detections are removed, as these may influence the classifier training in a negative way. This results in 29 different subsign classes, where frequently occurring texts are included separately, while less frequent texts are covered by a general subsign-with-text class (in practice, the text itself is unconstrained). As non-subsign detections may occur, we have also included a non-subsign class. Figure 4 displays the majority of the subsign categories within our training set.

4 EXPERIMENTS AND RESULTS

4.1 Subsign Detection Dataset Description

The above-described system is evaluated on a large and diverse test set, not overlapping with any of the used training sets. This set is constructed by applying our road sign recognition system (Hazelhoff et al., 2012) on a moderate-size geographical area, containing rural roads, smaller towns and a part of a city environment. Within this region, street-level panoramic images are captured on each public road, using an inter-capturing interval of 5 meter. This set contains $40,128$ images (about 200 km of road) in total. As these images are captured from driving vehicles under various weather and lighting conditions, cover various environments and represent a geographical area, we consider this test set as a representative and real-world dataset.

Within this region, $3,104$ $3D$ signs are identified by our traffic-sign inventory system (Hazelhoff et al., 2012), where each sign is detected in about 5 different images. These signs are employed to assess the detection and classification performance of our subsign recognition approach. From the $3,104$ signs, 397 (about 12.8%) have a subsign, where we should note that although this may look like a low percentage, this corresponds to the real-world occurrence of subsigns. Furthermore, we should note that the signs are captured from various viewpoints and distances (and thus with various resolutions), and are captured under various lighting and weather conditions. For each of the $3D$ signs, both the presence of a subsign and its optional subsign code are marked manually as ground truth. Example images of detected signs and regions

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Figure 5: Example images contained in our test dataset.

below these signs are shown in Fig. 5.

We have assessed the detection and classification accuracy separately, as both aim at different aspects (i.e. *the occurrence of a subsign* and *its specific type*).

4.2 Detection Performance

The detection performance is assessed for both the single-image detection and the multiview detection stages. During evaluation, the following performance metrics are employed:

- Precision: $\frac{TP}{TP+FP}$,
- Recall: $\frac{TP}{TP+FN}$,

where TP denotes the correct detections (true positives), FP denotes the erroneously found subsigns (false positives) and FN the missed subsigns (false negatives). Fig. 6 displays the recall-precision curves for both the single-image and multiview detection stages. For clarity, we additionally included a zoomed-in version, focusing at the high recall region in Fig. 7. We should note that the sharp bend in the curve for the single-image detection occurs because for about 10% of the images, the subsign is completely invisible (e.g. due to occlusions).

It can be observed that multiview detection significantly outperforms the single-image detection. This is explained by the fact that many of the contained signs are captured from a relatively large distance, from non-ideal viewpoints, or are (partially) occluded, complicating subsign detection. Moreover, the single-image detection stage is unable to retrieve about 11% of the present subsigns, e.g. due to aforementioned reasons. When combining the different detections of the same sign, the optimal views can be exploited, which increases the detection performance

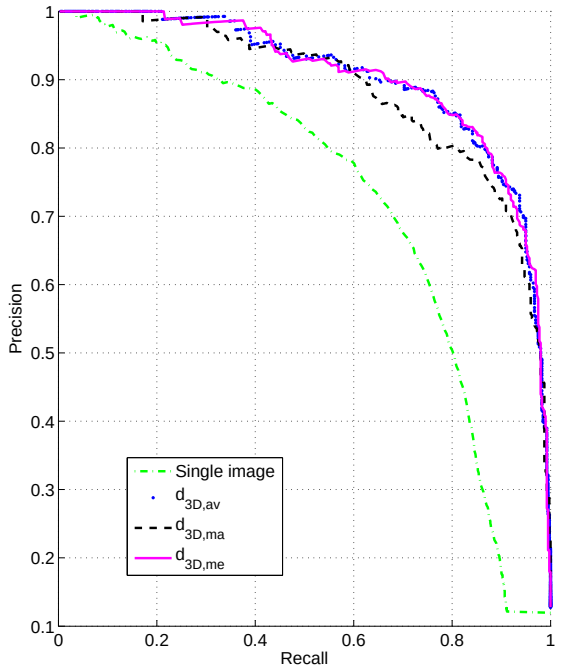

Figure 6: Recall-precision curves for both the single image and multiview detection stages.

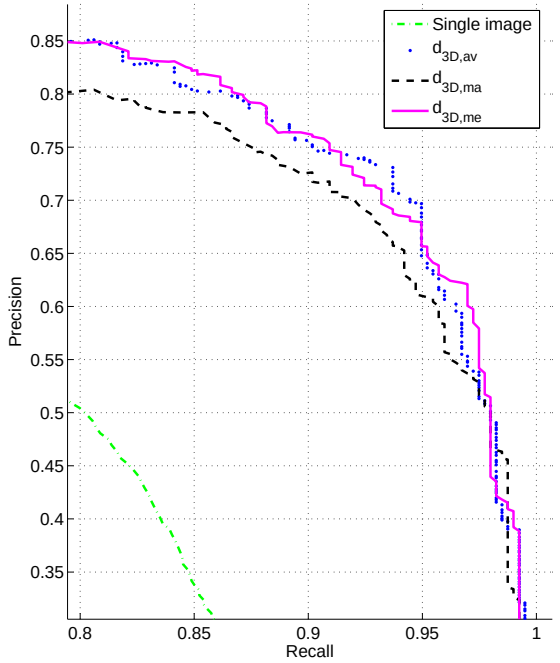

Figure 7: Recall-precision curve zoomed-in around the high recall region.

Table 1: Summary of the classification results. The last two rows are subclasses of types known to our classifier.

	Total		Correctly classified		Found as non-subsign	
Total detections	757	100.0%	-	-	-	-
Non-subsign	368	48.6%	0	0%	63	17.1%
Unknown types	17	2.2%	0	0%	0	0%
Known types	372	49.1%	340	91.4%	0	0%
-common subsigns	*200*	*53.8%*	*178*	*89.0%*	*0*	*0%*
-text subsigns	*172*	*46.2%*	*162*	*94.2%*	*0*	*0%*

considerably. We have found that for a high recall, the three multiview methods behave similarly, where the *median* and *average* methods outperform the *max* method for the lower recall region. Numerically, at a detection rate of 98%, the correct detection score is 51.3%, resulting in 757 signs out of the $3,104$ total 3*D* signs marked as containing a subsign. Although a precision of 51.3% looks quite low, subsigns only occur in a minority of cases, implying that for inclusion of subsigns in the inventory result, only about 24.1% of the 3*D* signs have to be checked, thereby reducing the amount of manual checks with a factor 4.1 compared to evaluating all 3*D* signs by hand.

Considering the processing time, the single-image detection stage takes ~ 1.9 seconds per detection, resulting in an average processing time of about 9.5 seconds for each 3*D* sign. We should note that although this looks rather slow, this computation time applies to each 3*D* sign (not each image) and is valid for a single-threaded implementation measured on a 2009 i7-920 CPU, operating at 2.67 GHz. Also, we should point out that the subsign detection stage adds about 0.9% of the total computational load required for all other components within the road-sign inventory system. Since this task can be performed in parallel for each 3*D* sign (employing multi-core computers in a distributed computing environment), this processing can be distributed to reduce the elapsed time of this stage to a sufficiently low number suitable for our purpose.

4.3 Localization and Classification Performance

After multiview detection, each 3*D* signs identified as containing a subsign is subject to classification. We consider three subgroups, as indicated in Table 1. The falsely detected subsigns are grouped into the non-subsign category, which may be classified as a non-subsign (an additional class in our classification system) but can never be classified correctly. The second category contains the minority of subsign types that are not known to our classifier (and are also non-text), and are therefore always classified incorrectly. The third group contains the types known by our classifier. This category is divided into two segments: common subsigns and text subsigns. The first category contains all subsigns that occur frequently, including broadly used text subsigns, which are all learned as a separate class by our classifier. The second category contains all other text subsigns, as in practice the text can be unconstrained and may even differ in a single letter. The classification results are summarized in Table 1. As can be noticed, 91.4% of the subsigns with a type known to our classifier is recognized correctly. Furthermore, 17.1% of the falsely detected subsigns is recognized accordingly, decreasing the number of false positives and resulting in a correct detection ratio of about 56% (at a detection rate of 98%).

When evaluating the processing time, the localization and classification stages require 2.3 seconds per 3*D* sign, which results in an average total processing time slightly above 10 seconds per 3*D* sign for the complete subsign recognition system. This adds about 0.9% to the total computational load required for all other components within the road sign inventory system.

5 CONCLUSIONS AND FUTURE WORK

This paper has presented a system for recognition of subsigns as an extension for automated traffic-sign recognition systems. The system first analyzes of the regions below each detected traffic sign (given by the road sign recognition system) using generic object detection methods. Second, the consecutive detections of the same 3*D* sign are combined to determine the presence of a subsign for each 3*D* sign. For signs identified as having a complimentary sign, the subsign pixel region is determined for each detection, and these regions are then subject to classification to retrieve the subsign type.

Performance evaluation on a large and diverse test set showed that reliable detection is feasible, where at a detection rate of 98%, about 51% of the detections is correct. Since subsigns are sparse (i.e. only

occur below the minority of signs), this results in a significant reduction in manual effort to include subsigns in road-sign inventories. This score may be increased by using the statistics of combinations of primary sign types and subsign occurrences, to exclude certain sign types from subsign detection. Classification of subsigns is challenging, since subsigns are small and have varying resolutions and a high diversity in contents, where falsely detected subsigns are also present. Nevertheless, 91.4% of the subsigns with a class known to our classifier are classified correctly. Additionally, 17.1% of the falsely detected subsigns are recognized accordingly as non-subsign. This implies that by a combined detection and classification approach, the correct detection ratio is further improved to $\sim 56\%$ for a detection ratio of 98%.

REFERENCES

Hamdoun, O., Bargeton, A., Moutarde, F., Bradai, B., and Chanussot, L. (2008). Detection and recognition of end-of-speed-limit and supplementary signs for improved european speed limit support. In *15th World Congress on Intelligent Transport Systems (ITS)*, New York, États-Unis.

Hazelhoff, L., Creusen, I. M., and De With, P. H. N. (2012). Robust detection, classification and positioning of traffic signs from street-level panoramic images for inventory purposes. In *Applications of Computer Vision (WACV), Workshop on*, pages 313 –320.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision (IJCV)*, 60(2).

Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H., and Lopez-Ferreras, F. (2007). Road-sign detection and recognition based on support vector machines. *Intelligent Transportation Systems, IEEE Transactions on*, 8(2):264 –278.

Maldonado-Bascon, S., Lafuente-Arroyo, S., Siegmann, P., Gomez-Moreno, H., and Acevedo-Rodriguez, F. (2008). Traffic sign recognition system for inventory purposes. In *Intelligent Vehicles Symposium, 2008 IEEE*, pages 590 –595.

Nienhuser, D., Gumpp, T., Zollner, J., and Natroshvili, K. (2010). Fast and reliable recognition of supplementary traffic signs. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, pages 896–901.

Overett, G. and Petersson, L. (2011). Large scale sign detection using hog feature variants. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 326 –331.

Puthon, A.-S., Moutarde, F., and Nashashibi, F. (2012). Subsign detection with region-growing from contrasted seeds. In *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, pages 969–974.

Timofte, R., Zimmermann, K., and Van Gool, L. (2009). Multi-view traffic sign detection, recognition, and 3d localisation. In *Applications of Computer Vision (WACV), 2009 Workshop on*, pages 1 –8.

Timofte, R., Zimmermann, K., and Van Gool, L. (2011). Multi-view traffic sign detection, recognition, and 3d localisation. *Machine Vision and Applications*.